

## Конверсия карточного каталога книг в электронную форму: опыт зональной научной библиотеки УГТУ-УПИ

Только наличие электронного каталога библиотеки, в котором отражен весь фонд, позволяет внедрить информационные процессы в работу библиотеки с возможностью получения преимуществ, перевешивающих затраты на информатизацию. Без полного электронного каталога (ЭК) невозможно дальнейшее продвижение по основным направлениям информатизации библиотеки, в первую очередь, автоматизации процессов, связанных с книговыдачей, контролем и учетом фонда и, в значительной степени, комплектованием.

Процессом перевода в электронную форму собственных каталогов занимались и занимаются множество библиотек. Крупнейшие западные библиотеки начали этот процесс еще в 70-е гг.; в наиболее значительных библиотеках России конверсия была осуществлена в конце 90-х. Необходимо отметить, что в крупных западных вузовских библиотеках сплошная конверсия каталогов предшествовала информатизации. В России вследствие бедности библиотек все процессы начались с деления фонда на новые книги, уже описанные в электронном виде, и старые, отраженные только в традиционном карточном каталоге. Затем постепенно и по этапам ретроспективная часть фонда стала конвертироваться в электронную форму. Часть каталога за предыдущие годы была введена и в зональной научной библиотеке Уральского государственного технического университета УПИ (ЗНБ УГТУ-УПИ). Идея при этом состояла в том, чтобы преобразовать в электронную форму законченные блоки информации, например все книги по металлургии, все учебники, фонды некоторых отделов и т.п. Однако это не решало основной задачи – полного ввода каталога с возможностью дальнейшего поиска. Кроме того, темпы ввода, обусловленные ограниченностью основного ресурса, измеряемого человеко-часами работы, никак не могли нас устроить. При существующей скорости ввода это была бы задача не на одно десятилетие.

Как известно, существует два основных подхода к проведению конверсии каталога в машиночитаемую форму. Один подход заключается в том, чтобы вводить информацию вручную с каталожных карточек или напрямую с книг. Ввод информации с книг характерен в основном для библиотек, обладающих небольшим и легкодоступным фондом. При этом ускорить работу можно, привлекая дополнительную рабочую силу (студентов) и используя сверхурочную работу сотрудников, а также заимствуя записи из баз данных РКП, МГУ и других библиотек. Выигрыш такого подхода – наличие полных описаний книг, отсутствие ошибок при распознавании, возможность параллельной редакции каталогов. Однако скорость ввода при этом минимальна: от 5 до 10 карточек, обрабатываемых одним опытным сотрудником в час. Привлечение сторонней рабочей силы приводит к повышению количества ошибок при вводе, на выявление и редакцию которых также уходит много времени. Для того чтобы обеспечить ввод ключевых слов и предметных рубрик, в некоторых библиотеках приходится использовать предварительную обработку карточек опытным каталогизатором или библиографом. Заимствованные записи дорабатываются для приведения в соответствие с собственными методическими рекомендациями, что зачастую требует не меньших трудозатрат, чем ввод новых записей. Кроме того, данный вариант можно оправдать только лишь сокращением числа ошибок при вводе, хотя ни один электронный каталог, в том числе РКП или библиотеки Конгресса США, не идеален и содержит ошибки и опечатки. Для поблочного ввода информации этот путь в библиотеке УГТУ-УПИ использовался, однако для полной конверсии каталога он явно не годился.

Другой подход заключается в том, чтобы сканировать карточный каталог, распознавая карточки, и на основе распознанной информации автоматически создавать записи электронного каталога. Этим путем пошло большинство крупнейших библиотек России. Наиболее яркий пример – Российская государственная библиотека [3]. Таким способом обеспечивается высокая скорость процесса конверсии; а за счет того, что работа, как правило, осуществляется сторонней организацией, поскольку библиотеки не располагают необходимым оборудованием, высвобождается рабочее время сотрудников библиотеки. В нашей стране существует несколько организаций, выпол-

няющих подобного рода работы – Прософт-М, Гипер и др. Главным недостатком такого сотрудничества является высокая стоимость работы, что для вузовской библиотеки является существенной проблемой, так как ей очень сложно изыскать необходимую сумму. Несмотря на то, что качество сканирования карточек довольно высокое, вместе с тем вероятность ошибок остается, и записи все равно придется редактировать. Кроме того, зачастую на карточках информация расположена нестандартным образом или обнаруживается ее недостаток, имеются зачеркнутые инвентарные номера, индексы и другие пометки библиотекарей. Все это требует не столько качественного ввода, сколько отработанного процесса верификации введенной записи в соответствии с реалиями библиотеки-фондодержателя. В силу данных причин метод сканирования в вузовских библиотеках имеет не столь массовое распространение [2].

При разработке концепции конверсии каталога в нашей библиотеке основное внимание было уделено созданию и отработке процесса, обеспечивающего следующие характеристики:

- разделение всего процесса на последовательность регламентированных этапов, поддающихся контролю и нормированию;
- каждый этап состоит из простых технологических операций; наиболее трудоемкие и длительные этапы допускают использование неквалифицированной рабочей силы без снижения качества и скорости ввода;
- результат каждого этапа проходит двойной контроль: при приеме работы в автоматическом режиме (отсутствие незаполненных обязательных полей и т.д.) и на каждом из последующих этапов с возможностью вернуться назад для исправления ошибок;
- все этапы, за исключением последнего (сверки введенных записей с книгами *de visu*), допускают параллельное исполнение, причем не обязательно на оборудовании библиотеки, что позволяет достичь скорости ввода записей;
- каждая из введенных записей должна быть доступна для поиска сразу по завершении этапа обработки, без ожидания результата проверки и окончания следующих этапов;
- в любой момент времени должна быть известна информация о выполненном и предстоящем объеме работ.

Основное требование к создаваемому полному электронному каталогу библиотеки УГТУ-УПИ заключается в обеспечении возможности многоаспектного поиска информации, превосходящего поиск по всем имеющимся традиционным карточным алфавитным и систематическим каталогам.

Процесс поиска книги читателем заключается в просмотре им традиционных карточек или электронных записей, при котором определяющую роль играет знание предмета поиска, а информация о книге является сигнальной для окончательного определения релевантности результата критерию поиска и заказа книги в отделе обслуживания. Для правильной идентификации книги и получения информации, по которой можно заказать книгу или составить правильную библиографическую ссылку, достаточно информации, представленной на печатных библиографических карточках, выпускавшихся Книжной палатой. Таким образом, оправдана возможность представить читателю в качестве библиографического описания отсканированное изображение карточки. Такое изображение без дополнительной обработки может служить также для восстановления утраченных или утерянных карточек каталога.

На наш взгляд, для проверки наличия книги в библиотеке читателю достаточно найти ее в каталоге по унифицированному заглавию и авторам, что подтверждено долгой и успешной эксплуатацией алфавитного каталога [4]. Для повышения точности поиска достаточно ввести заглавие и год издания книги. Для обеспечения предметного поиска достаточно индексов УДК или ББК, уже проставленных на карточке при индексации книги в момент ее первичного описания. Возможная недостаточность глубины индексов универсальных классификаций компенсируется использованием локальных рубрикаторов, более полно отражающих информационные потребности организации [1]. Поскольку перечень индексов локального рубрикатора четко определен, то появляется возможность контролировать ошибки при вводе и полуавтоматически проверять правильность индексации с привлечением систематизатора. Как заметил Э.Р. Сукиасян [5], ключевые слова, поиск по которым, как правило, является основным тематическим поиском в электронных каталогах, будучи одним из самых трудоемких элементов ввода записей (вернее, требующих специальной подготовки и опыта работника), не являются обязательным элементом описания книги. Кроме того, ключевые слова не существуют в готовом виде на карточке, а для того, чтобы снаб-



Скорость сканирования – примерно 6-7 часов чистого времени для обработки содержимого одного ящика (в среднем – около 600 карточек) двумя библиотекарями (один работает за компьютером, другой собирает и разбирает планшеты с карточками). Для улучшения качества сканирования карточки данного каталога необходимо было разбирать по цвету, а затем восстанавливать алфавит, что явилось дополнительной задержкой. В ГАК темных неконтрастных карточек – примерно 1,5-2% от общей массы, поэтому в дальнейшем можно будет сканировать, не нарушая алфавита, что значительно ускорит дело – до 4 часов на ящик. Сканирование производилось силами привлеченных студентов. Это вызвало некоторые проблемы. Незаинтересованность студентов в конечном результате приводит к тому, что без постоянного контроля за их работой качество ввода резко падает. Впоследствии возникла необходимость повторного сканирования более 10% карточек. Однако, заранее предполагая такие результаты, мы сознательно пошли на привлечение студентов к работе по сканированию, а затем вводу и проверке информации, так как для конверсии всего ГАК, где объем работы будет на два порядка больше, использование сторонней рабочей силы – необходимое условие.

Далее отсканированные изображения блока карточек в формате tif были автоматически разрезаны программой, написанной на языке Perl с использованием свободно распространяемой библиотеки ImageMagick. В итоге было получено 6549 картинок в формате tif, которые затем были преобразованы в формат djvu, специально предназначенный для представления в Интернет отсканированных изображений текста. При этом преобразовании достигается значительная экономия дискового пространства (более чем наполовину).

По результатам эксперимента для дальнейшей работы по сканированию карточек ГАК планируется применить следующие усовершенствования данной технологии: использовать для сканирования протяжной сканер, исключая процесс последующей резки изображений, что снизит ошибки при сканировании (таких как обрезка краев, перевернутая карточка, несовпадение цветности) и обеспечит большую скорость сканирования (на порядок); сканирование в соответствии с ящиками каталога и без нарушения алфавита с последующей проверкой качества; дальнейшее редактирование отбракованных отсканированных карточек без дополнительных перемещений ящиков каталога. В дальнейшем карточки с отметкой ЭК, т.е. уже введенные в электронный каталог, не будут сканироваться.

В июне 2003 г. силами студентов по сканированным изображениям были введены в БД краткие описания книг. Это делалось с использованием вышеописанной программы ввода (рис. 2).

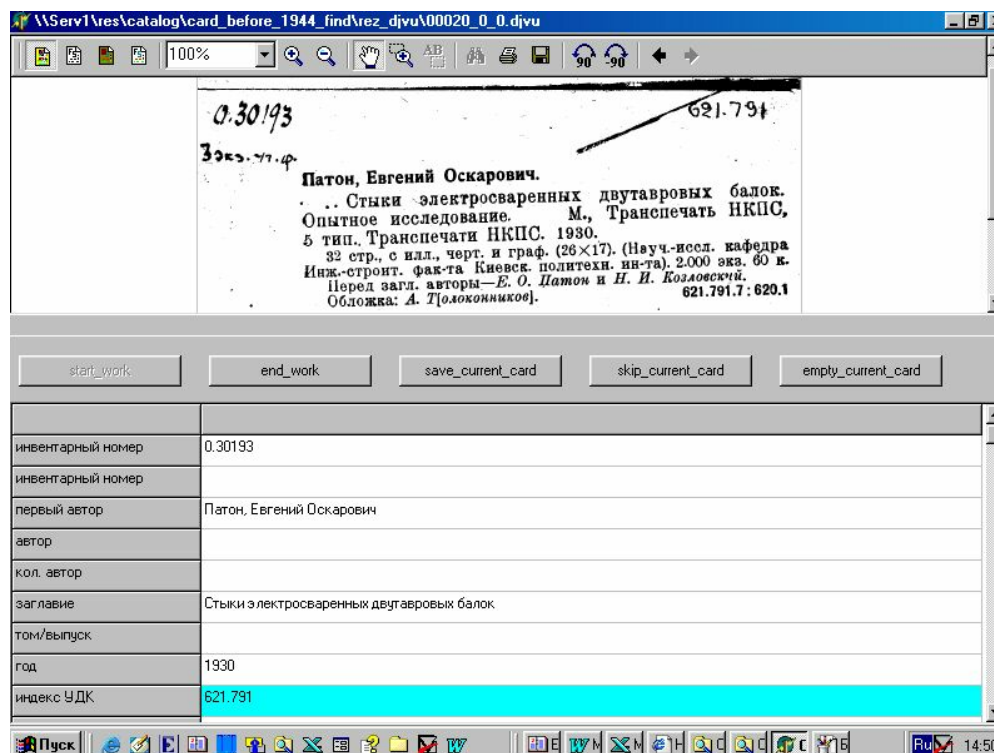


Рис. 2. Изображение рабочего окна программы ввода

Если программа при вводе обнаруживала, что вводимый инвентарный номер уже существует в БД, то сообщение об этом появлялось на экране и запись не вводилась.

Надо заметить, что скорость ввода зависит от умения студента работать с клавиатурой. Максимальная скорость ввода – до 50 карточек в час. Реально средний студент за сокращенный рабочий день (четыре часа) набивал около 150 карточек. Указанная скорость не является чем-то исключительным: существует пример ввода библиографических описаний со скоростью 300 записей за шестичасовой рабочий день [2]. Большого ускорения ввода можно достичь путем оптимизации интерфейса и использования операторов, владеющих методом слепой печати. Для контроля качества ввода студентам предлагается выдавать проверочные задания по вводу информации с части карточек с дальнейшим определением рейтинга операторов и более пристальной проверкой нерадивых. Предполагается также ввести отдельный этап для проверки дублетности описаний квалифицированным сотрудником, который будет по горячим следам проверять все случаи совпадения номеров, определяя необходимость повторного ввода информации с карточки или, при подтверждении дублетности, приписывать изображение к существующей записи ЭК.

По окончании процесса был распечатан список записей на книги до 1946 года издания, включающий более чем 5,7 тыс. позиций. Меньшее, по сравнению с исходным количеством карточек, число записей объясняется, во-первых, наличием добавочных карточек на второго автора в каталоге и, во-вторых, некоторым количеством карточек очень плохого качества, которые не были использованы для описания (не читались). Силами студентов под присмотром работников отдела книгохранения этот список был сверен с книгами, стоящими на полках в хранилище и отделах. Снятые с полок книги были просмотрены дважды: сначала по электронному каталогу на предмет наличия в БД – в итоге около 100 книг, не найденных ранее в списке по разным причинам (ошибки студентов, опечатки в списке), были найдены и возвращены в хранилище. Также было выявлено 400 книг, номера которых не соответствуют тем номерам, которые значатся в БД (опечатки и ошибки ввода, а также ошибки, накопившиеся в печатном каталоге). Затем книги были еще раз сверены с печатным каталогом на предмет выявления тех книг, карточки на которые были плохо отсканированы или некачественно введены записи. Было выявлено еще 568 книг, описания которых имелись в каталоге. Карточки на них были отсканированы заново и введены дополнительно в БД. Кроме того, выявилось также некоторое количество книг, карточки на которые по тем или иным причинам были утрачены и отсутствовали в карточном каталоге.

На каждый из получившихся массивов книг (с ошибками в номерах, дополнительно введенные и не значащиеся в каталогах) были также подготовлены списки, выстроенные по инвентарным номерам. Все списки (исходный и три новых) были сверены в отделе комплектования с инвентарными книгами на предмет выявления недостающих книг (все еще не введенных), а также ранее ошибочно списанных. Таких книг было обнаружено около 50 экземпляров. Все книги, электронные записи на которые по итогам проверки отсутствовали, будут введены в основной электронный каталог, как если бы это были новые поступления.

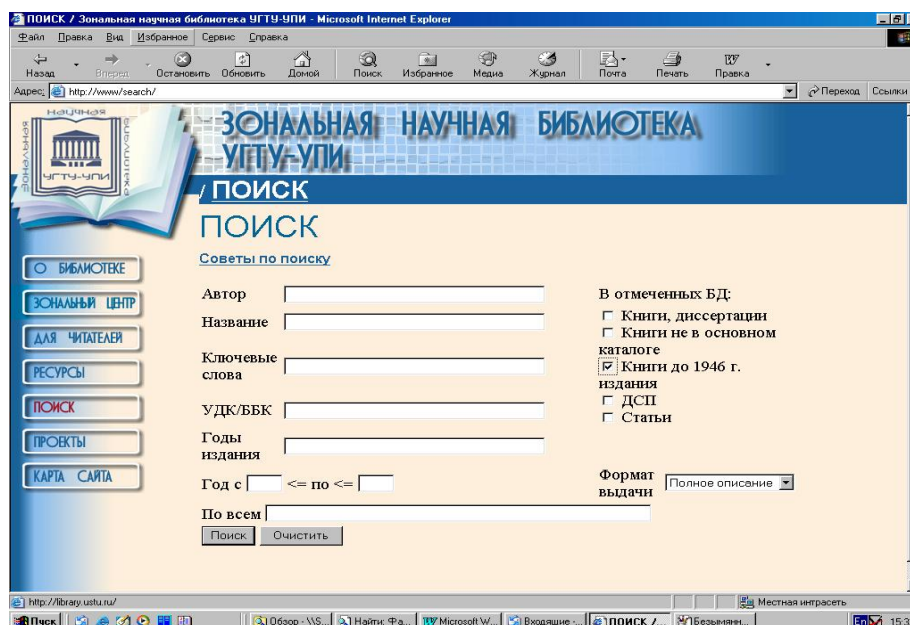


Рис. 3. Поисковая страница сайта библиотеки УГТУ-УПИ

Для завершения процесса конверсии необходимо также проиндексировать описания книг, в которых УДК или ББК отсутствуют, или не соответствуют современному рубрикатору электронного каталога нашей библиотеки.

Описания книг в БД имеют связь с исходными изображениями. При поиске на сайте нашей библиотеки имеется возможность выбора данной БД наряду с другими частями электронного справочного аппарата библиотеки (рис. 3).

В результате поиска по БД книг до 1946 года издания пользователь получает краткое описание книги (рис. 4) и, выбрав ссылку «Карточка каталога», может посмотреть исходное изображение карточки (рис. 5).

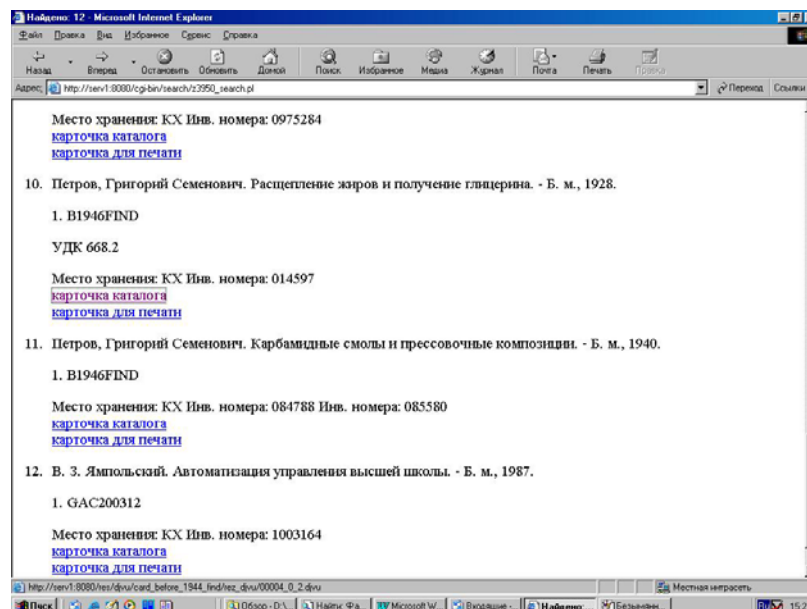


Рис. 4. Выдача результатов поиска в каталоге книг до 1946 года издания

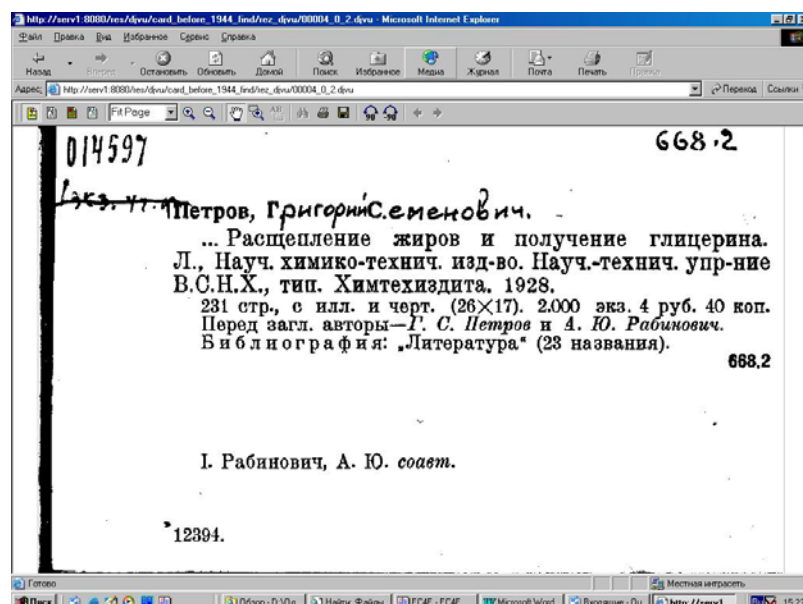


Рис. 5. Изображение отсканированной карточки для читателя

Для тематического поиска книг, введенных в электронный каталог подобным образом, будет использован рубрикатор электронного каталога ЗНБ УГТУ-УПИ, идеология создания которого была описана ранее [1].

В настоящее время продолжается работа по сканированию карточек основной части ГАК и вводу записей в базу данных. Предварительно эти записи будут автоматически проверены на возможность заимствования библиографического описания из каталогов РКП и других баз данных.



По завершении ввода информации с карточек каждого отдельного ящика будет формироваться объединенный список отсканированных и введенных в ЭК записей в порядке алфавита для сверки полноты ввода, при этом найденные карточки будут либо сканироваться вновь, либо информацию с них введут вручную. Одновременно сверенный список можно использовать для проверки читательского алфавитного каталога с допечаткой недостающих карточек. Наличие изображений карточек также позволит при необходимости пополнять или восстанавливать карточки традиционных каталогов подсобных фондов отделов.

После завершения работы по вводу могут быть сформированы списки по индексам рубрикатора в целях сверки их соответствия индексам на книгах, а также сверки наполнения отдельных рубрик систематического или предметного каталогов.

Мы надеемся, что с привлечением соответствующих ресурсов (студентов, новой сканирующей техники) процесс полной конверсии каталога будет в основном завершен к концу 2005 г., а с учетом проверки результатов – к началу 2006/2007 учебного года.

### **Библиографический список**

1. Акоев М.А. Систематический рубрикатор электронного каталога / М.А. Акоев, О.Г. Васильев // Университетская библиотека: проблемы и решения: Труды Пятой Всерос. науч.-практ. конф. Екатеринбург, 2002. С. 96-100.
2. Воройский Ф.С. Организация и технология переработки карточных каталогов в машиночитаемую форму для создания электронных каталогов / Ф.С. Воройский // Науч. и техн. б-ки. 1999. №1. С. 106-116.
3. Лавренова О.Л. Организация и технология ретроспективной конверсии каталогов / О.Л. Лавренова, М.-Э. Фреон // Электрон. б-ки. 2000. Т. 3, вып. 3. <http://www.iis.ru/el-bib/2000/200005/LF/lfr.html>
4. Сукиасян Э. Р. Библиотечные каталоги: Метод. материалы / Э.Р. Сукиасян. М.: Профиздат, 2001. 192 с.
5. Сукиасян Э.Р. Пришло ли время «закрывать» систематический каталог? / Э.Р. Сукиасян // Науч. и техн. б-ки. 2002. №12. С. 18-30.

**Л. М. Лошкарева, Г.С. Щербинина**  
*УГТУ-УПИ, Екатеринбург*

### **Проблемы многоязычия в каталогизации**

Современное развитие процессов автоматизации и Интернета, с одной стороны, открывают неограниченные возможности для знакомства с мировыми информационными ресурсами, с другой стороны, остро ставят проблему ориентации в них, поскольку проблема многоязычия является для пользователей и каталогизаторов одной из самых непростых.

Предметом данной статьи являются проблемы многоязычия в каталогизации четырех категорий документов, каждая из которых имеет собственные приемы описания. Кроме того, необходимо обозначить общие тенденции в области каталогизации подобных документов в современных условиях. Итак, к первой категории относятся документы на иностранных языках, изданные за рубежом и имеющие статус иностранных. Это те документы, которые библиотеки получают по валютной подписке, по грантам, в качестве даров от читателей и т.д. Ко второй категории могут быть отнесены документы российских авторов, опубликованные на иностранном языке, но изданные в России, например препринты, а также книги для чтения, которые, в свою очередь, могут принадлежать перу как российских, так и зарубежных авторов. К данной категории можно отнести адаптированные тексты зарубежных авторов в помощь изучению иностранных языков. К третьей категории относятся документы, переведенные с иностранного языка на русский. И к четвертой категории относятся издания на русском языке, имеющие в заглавии слова в иной, как правило, латинской, графике.

Необходимо отметить, что проблемы, связанные с каталогизацией документов обозначенных категорий, касаются области языка библиографической записи в целом и ее отдельных частей: выбора языка описания, языка заглавия и языка содержательной каталогизации. В данном